

Correlação e coeficiente de correlação linear de Pearson

Professor Fiore

Em estatística, o **coeficiente de correlação** é um número que indica o grau com que duas variáveis se relacionam. Por exemplo, em uma linha de montagem, o custo total de produção pode estar correlacionado linearmente com a quantidade produzida. A população de uma cidade pode estar correlacionada com o tempo de forma exponencial. A quantidade de remédios que uma pessoa toma pode estar correlacionada com a idade da pessoa.

Uma boa forma de verificar qual tipo de correlação existe entre os dados é por observar o diagrama de dispersão, gráfico com pontos dispersos no plano cartesiano. Ao fazer isso, lembre-se que existem várias formas de correlacionar dados, como funções polinomiais, exponenciais, logarítmicas, trigonométricas, entre outras.

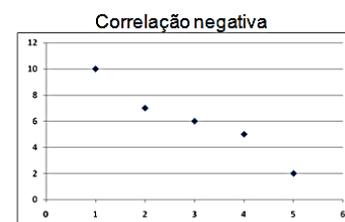
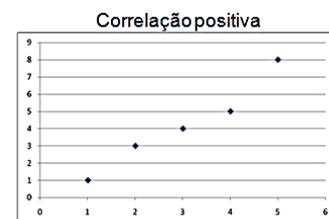
Um bom coeficiente de correlação pode **indicar** uma possível **relação de causa e efeito** entre duas grandezas, mas **não pode garantir** essa relação. Por exemplo, em um grupo de pessoas, o número de cafezinhos que cada uma toma pode estar correlacionado linearmente com a quantidade de sobrinhos que cada um tem, mas isto não garante relação de causa e efeito entre os dados.

Em muitos casos os dados estão relacionados linearmente, ou seja, eles seguem uma **tendência linear** e neste caso podemos usar o **coeficiente de correlação linear de Pearson** (r). Para calculá-lo, entre as variáveis numéricas x e y . A fórmula abaixo indica o valor do coeficiente linear de Pearson, onde n é o número de pares de valores.

$$r = \frac{n \cdot \sum x_i \cdot y_i - (\sum x_i) \cdot (\sum y_i)}{\sqrt{(n \cdot \sum x_i^2 - (\sum x_i)^2) \cdot (n \cdot \sum y_i^2 - (\sum y_i)^2)}}$$

Sobre o valor do coeficiente é possível afirmar que:

- Ele sempre estará entre -1 e 1. Ou seja: $-1 \leq r \leq +1$.
- Valores positivos indicam uma correlação crescente.
- Valores negativos indicam uma correlação decrescente.
- Quando o valor de r é 1 ou -1, há uma correlação linear perfeita.
- Próximo de zero, a correlação linear é praticamente inexistente.
 - Próximo -0,25 ou 0,25 a correlação é muito fraca.
 - Próximo de -0,50 ou 0,50 a correlação é fraca/média.
 - Próximo de -0,75 ou 0,75 a correlação é média/boa.
 - Próximo de -0,90 ou 0,90, a correlação é forte.



Quando a correlação é positiva o aumento no valor de uma variável implica no aumento do valor da outra. E para correlações negativas, o aumento no valor de uma variável implica na redução do valor da outra.